Search (Non-)Neutrality and Impact on Innovation

Pierre L'Ecuyer, Patrick Maillé, Nicolás Stier-Moses, Bruno Tuffin

Search (Non-)Neutrality and Impact on Innovation

Pierre L'Ecuyer
Université de Montréal
C.P. 6128, Succ. Centre-Ville,
Montréal, H3C 3J7, Canada
lecuyer@iro.umontreal.ca

Patrick Maillé
Telecom Bretagne
2, rue de la Châtaigneraie
35576 Cesson Sévigné
Cedex, France
patrick.maille@telecom-bretagne.eu

Nicolás Stier-Moses
Universidad Torcuato Di Tella
Saenz Valiente 1010, Buenos
Aires, Argentina
nstier@utdt.edu

Bruno Tuffin
Inria
Campus Universitaire de
Beaulieu, 35042 Rennes
Cedex, France
bruno.tuffin@inria.fr

ABSTRACT

The search neutrality debate, as a parallel to the network neutrality debate, is raging worldwide, with search engines accused of biasing the ranking of their organic links to provide a competitive advantage to their own content. In a recent paper, we have designed a model and determined the optimal ranking policy for a search engine as a trade-off between short-term revenue (based on the potential immediate gain from high-ranked links) and long-term revenue (based on the satisfaction of users due to the relevance of the ranking). We here apply this model to investigate on an example whether non-neutrality impacts innovation. We illustrate that a revenue-oriented search engine may indeed deter innovation at the content level, hence the validity of the argument (without necessarily meaning that search engines *should* be regulated).

Categories and Subject Descriptors

C.2.3 [**Computer Systems Organization**]: Computer-Communication Networks—*Network Operations*

General Terms

Economics, Neutrality, Search Engines, Vertical Integration, Competition

1. INTRODUCTION

In recent years, there has been a debate about *search neutrality*, relating it to the more well-known *network neutrality debate* which was more focusing on Internet service providers [6]. Indeed, some search engines (SEs) have been under scrutiny by individuals and organizations that oversee the Internet, as well as by regulators in various countries, because some believe that the organic search ranking is not based only on objective measures of relevance, but also accounts for some revenue-making ingredients [2]. For example, it has been said that Google may favor YouTube and other of its own content because of the extra revenue it generates. It can be direct revenue due to sold products, but more often indirect revenues from advertisements due to visits to free web sites or applications. This bias in ranking has been experimentally illustrated, see [4, 6, 9] where it is shown for example that Microsoft (resp. Google) content is 26 times (resp. 17 times) more likely to be displayed on the first page of Bing (owned by Microsoft) (resp. Google) than on any other search engine. Search neutrality has been and is still discussed by governments and regulators, such as in the US by the Federal Trade Commission [1] and in a Senate hearing [8], or in Europe where Google is facing a \$6 billion fine. The question is whether or not a search engine should or not base its ranking on relevance on links only, and whether or not a non-neutral engine would harm the Internet economy by reducing competition, or innovation by favoring the incumbents that are known to generate profits. Such a behavior would prevent new applications/content from being shown, and hence from becoming known and successful. The question relates to other policy debates regarding whether and how to regulate the Internet, the most prominent example being the already mentioned network neutrality debate [7].

We have designed in [5] a model of a search engine capturing the trade-off between long-term gains (due to large numbers of visits) thanks to a ranking based on relevance and short-term gains due to highly ranked own content. We have characterized the (simple) optimal ranking policy for a search engine and shown how to compute it. Our goal in the present paper is to apply this optimal ranking in order to answer the question of whether or not non-neutrality, that is, a ranking not based on relevance only, hurts content innovation and investment.

The remainder of the paper is organized as follows. In Section 2, we briefly recall the main characteristics of the model in [5] and the main results obtained on it. Section 3 then presents the specific model studying the impact of a non-neutral engine on content innovation and investment, assuming that the search engine implements the previously described revenue-maximizing ranking policy.

2. MODEL: NON-NEUTRAL AND NEUTRAL SE RANKINGS

The search engine model and the corresponding results (revenue-maximizing ranking policy) are described in full detail in [5]; here we summarize them to develop further the content innovation aspect.

Consider a search engine receiving requests. To a request y corresponds a number m of pages to be displayed. We denote by $\{1, \ldots, m\}$ the set of corresponding pages, and define for each page i ($1 \le i \le m$) the relevance r_i of this page for request y. A ranking is a choice of permutation $\pi(y) = (\pi(y,1), \ldots, \pi(y,m))$ for each request y. A "fair" or "neutral" SE should rank the pages in decreasing order of relevances.

A request can therefore be defined as $y = (m, r_1, \ldots, r_m)$. But requests arriving randomly according to a given distribution, we denote the random request as $Y = (M, R_1, \ldots, R_M)$ with random corresponding number of pages and relevances: y is a realization of Y.

Now assume that to each request, there is an associated gain g_i for the SE if link i is clicked, which can be due to a direct sale or to indirect gains from advertisements on the associated page for example. The click-through rate (CTR) of link i if at position k is commonly assumed to be separable into a position effect θ_k and a relevance effect $\psi(r_i)$, hence a CTR $\theta_k\psi(r_i)$ (assuming $\theta_1 > \theta_2 > \cdots > \theta_m$), so that the gain from being displayed at slot k is $\theta_k\psi(r_i)r_i$. A random request is then extended to $Y = (M, R_1, G_1, \ldots, R_M, G_M)$, with those gain values. A revenue-oriented SE would be tempted to sort links for any y according to the $\tilde{g}_i = \psi(r_i)g_i$, but this would not take into account the level of satisfaction of users who are more likely to reuse the SE if the ranking is relevant. A revenue-oriented SE actually would rather prefer to rank for each request in order to maximize

$$\lambda(r)(g + \beta)$$

where

- β is the average gain per visit due to *sponsored links*, that is how SEs are known to earn money
- $r = \mathbb{E}\left[\sum_{i=1}^{M} \theta_{\pi(Y,i)}\psi(R_i)R_i\right]$ is the average relevance over all requests depending on the choices of permutations
- $\lambda(r)$ is the average number of requests per unit of time which is assumed to increasingly depend on the quality (the average relevance) of the SE
- $g = \mathbb{E}\left[\sum_{i=1}^{M} \theta_{\pi(Y,i)}\psi(R_i)G_i\right]$ is the average gain from organic links for the given ranking policy $\pi(Y, \cdot)$.

Remark that it encompasses the case of a "neutral" SE ranking according to relevance by considering $G_i = 0\ \forall i$: indeed in that case the SE revenue is simply maximized by maximizing the average relevance r.

In [5], we have shown that, under conditions that we are "hiding" here for space reasons, the optimal ranking policy can be explicitly described. Denote by $\tilde{R}_i = \psi(R_i)R_i$ (resp. $\tilde{G}_i = \psi(R_i)G_i$) the relevance (resp. gain) from i weighed by its relevance-effect click probability. The optimal policy is simply, for any request Y, to rank links in a decreasing order of $\tilde{R}_i + \rho\tilde{G}_i$ for some real ρ. Such a policy is called a LO-ρ policy (Linear Ordering with weight ρ). It remarkably simplifies the search for an optimal policy to just finding the optimal parameter ρ, which can be done easily by simple optimization techniques.

3. IMPACT OF NON NEUTRALITY ON INNOVATION AND INVESTMENT

Our goal in this paper is to investigate the impact of a non-neutral SE, that is, of an SE applying the above optimal policy, on other actors, namely users and content providers. We especially aim at investigating whether or not a non-neutral SE harms competition and innovation. The numerical analysis we perform can be easily replaced with any other input data.

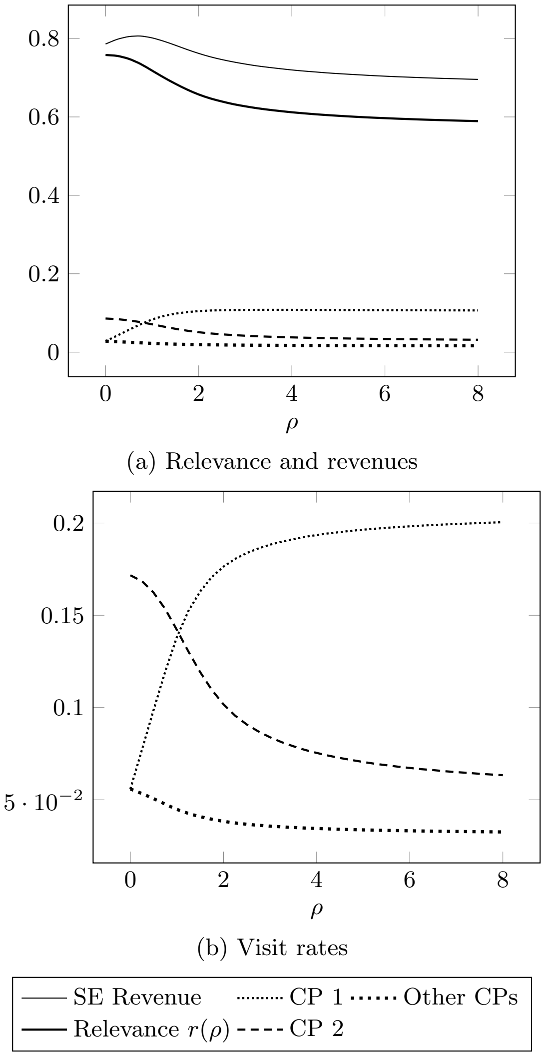
Consider that among the M pages corresponding to a given request, one exactly (say, page number 1 for simplicity) is served directly by the SE while the others are served by third parties. We then have $G_2 = \ldots = G_M = 0$. For our numerical illustrations, let us assume that always ten pages match a request ($M = 10$). In addition to the revenue coming from Page 1, the SE also receives an expected revenue of $\beta = 1$ per request from sponsored links. We furthermore assume CTR$(i) = \theta_i$ as specified in Table 1, values obtained from measurements in [3]. We also set $\lambda(r) = r$, and ψ to be the unit function (i.e., only position affect the click-through-rate). For $i = 1, 3, 4, ..., 10$, R_i and G_1 are all considered independent random variables uniformly distributed over $[0, 1]$. To consider the impact on innovation, we assume that CP 2 invests in quality and manages to improve its relevance distribution. More specifically, we assume that when it invests to reach a quality $z > 0$, the relevance of CP 2 becomes uniformly distributed over $[0, 1 + 20z]$ (instead of over $[0, 1]$). We still assume independence between R_2 and the other random variables (relevances and gain G_1). Additionally, since we want to consider the impact of a non-neutral ranking on the revenue of other content providers, we assume that they have too a gain uniformly distributed $[0, 1]$ and independent of the other random variables, however that gain is not to the SE so we still have $G_i = 0, i \ge 2$.

Figures 1(a) and 1(b) show simulation results when the SE ranks CPs according to $\tilde{R}_i + \rho\tilde{G}_i$, for varying values of ρ, and when $z = 2$. For a neutral ranking ($\rho = 0$), CP 2 logically makes more revenue than the other CPs, since it regularly gets higher ranking. However, when ρ increases and exceeds about 0.8, CP 1 becomes the one with highest revenue, despite its (stochastically) lower relevance. The optimal ranking for the SE is when $\rho \approx$0.75. This revenue-maximizing value is denoted by ρ^* (which depends on z), and we later assume that the SE will apply its LO-ρ^* ranking policy.

We now take the perspective of CP 2, and compute its optimal decision. CP 2 invests in quality z to modify its relevance distribution to $[0, 1 + 20z]$, anticipating that the SE is going to rank requests according to ρ^*. To simplify the example, we assume that an estimate of the distribution of Y is always immediately available to the SE (in real life there will be a delay to update the estimate when the distribution changes, which is fine if the distribution changes only slowly). The profit of CP 2 is the revenue from the search market, minus the unit investment cost times z. To optimize z, we simulated the outcomes for $z \in [0, 0.45]$. Figures 2(a) and 2(b) plot the resulting curves, with a unit cost of quality

Table 1: CTR values used in the simulations, taken from [3]

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
0.364	0.125	0.095	0.079	0.061	0.041	0.038	0.035	0.03	0.022



(a) Relevance and revenues

(b) Visit rates

Figure 1: Average relevance, revenues from the search market, and number of visits per unit of time for the case of vertical integration with investment

of 0.7. In both figures, we find that the difference between neutral and non-neutral revenues is very large for CP1. For CP2, the difference increases with the investment in quality, for two reasons: *i)* CP2 is impacted more frequently by the non-neutral ranking (being more often the best CP, it is more often artificially put behind CP1), and *ii)* CP1 also benefits from the attractiveness that CP2 creates, hence a large number of visits, and can afford to be even "less neutral" by increasing ρ^*. As a result, investments made by CP2 benefit for a very large part to CP1, and a bit to the other CPs due to the higher overall visit rate. But CP2 is significantly hurt by non-neutrality; and if it considers its optimal investment level we have:

- in a neutral regime, CP2 would select $z = 0.15$, and obtain a per time unit net revenue of 0.056;
- in a non-neutral regime, CP2 would select $z = 0.10$ (hence 33% less investment), and obtain a net revenue of 0.046 (hence a 18% decrease).

4. CONCLUSION

In this paper, we apply a model of revenue-maximizing ranking by a SE to compare it with a neutral ranking, based only on relevance. To investigate the impact of non-neutral ranking on content innovation, we have considered several CPs: one integrated by the SE (hence often favored in the rankings), one investing to improve its quality and revenues, and other independent ones. Our conclusion is that, as claimed by neutrality proponents, non-neutrality can significantly restrain innovation: for our example, innovation (investments in quality) are reduced by 33%, and the innovating CP makes 18% less than in a neutral regime. Of course, those conclusions strongly depend on the parameter values: with other values we have found that the impact of non-neutrality is not very large. Some deeper econometric studies would be necessary to find the most appropriate parameters, and our analysis could then be directly applied on those data.

5. REFERENCES

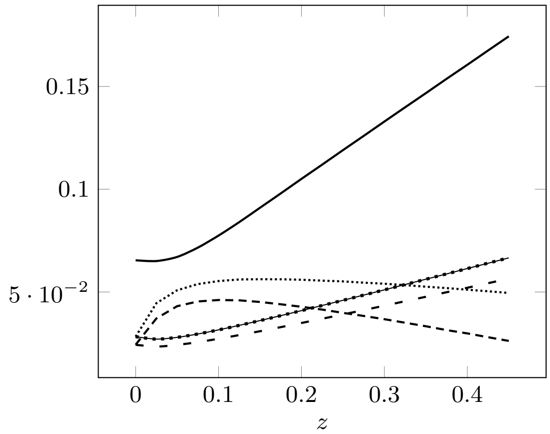
[1] J. Brill. Statement of the Commission regarding Google's search practices, 2013. `http://www.ftc.gov/public-statements/2013/01/statement-commission-regarding-googles-search-practices`. Last accessed Oct 2014.

[2] J. Crowcroft. Net neutrality: The technical side of the debate: A white paper. *ACM SIGCOMM Computer Communication Review*, 7(1), January 2007.

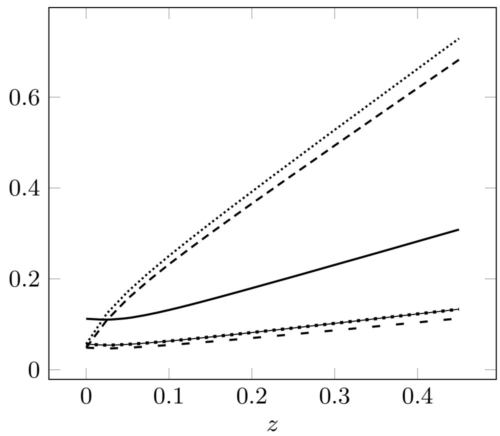
[3] R. Dejarnette. Click-through rate of top 10 search results in Google, 2012.

[4] B. Edelman and B. Lockwood. Measuring bias in "organic" web search. `http://www.benedelman.org/searchbias`, Jan 2011.

[5] P. L'Ecuyer, P. Maillé, N. Stier-Moses, and B. Tuffin. Revenue-Maximizing Rankings for Online Platforms



(a) CP revenues



(b) Visit rates

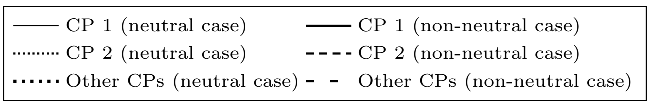


Figure 2: Revenues (including quality investment, at unit cost 0.7) and visit rates to various CPs as a function of the investment from CP 2

with Quality-Sensitive Consumers. Technical report, Inria, Feb. 2014. https://hal.inria.fr/hal-00953790.

[6] P. Maillé and B. Tuffin. *Telecommunication Network Economics: From Theory to Applications.* Cambridge University Press, 2014.

[7] A. Odlyzko. Network neutrality, search neutrality, and the never-ending conflict between efficiency and fairness in markets. *Review of Network Economics*, 8(1):40–60, 2009.

[8] D. Rushe. Eric Schmidt Google senate hearing – as it happened, 2012. `http://www.guardian.co.uk/technology/blog/2011/sep/21/eric-schmidt-google-senate-hearing` . Last accessed Oct 2014.

[9] J. D. Wright. Defining and measuring search bias: Some preliminary evidence. George Mason Law & Economics Research Paper 12-14, George Mason University School of Law, 2012.